DOCUMENT RESUME

ED 175 874 TM 007 669

AUTHOR Folsom, Ralph E., Jr.
TITLE Variance Components for NLS: Partitioning the Design Effect.
INSTITUTION Research Triangle Inst., Durham, N.C.
PUB DATE Jan 75
NOTE 25p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Variance; Cluster Analysis; Followup Studies; Graduate Surveys; High School Graduates; Longitudinal Studies; *Mathematical Models; *National Surveys; Probability; *Research Design; *Research Methodology; *Sampling; *Statistical Surveys; Weighted Scores
IDENTIFIERS *National Longitudinal Study High School Class 1972; Stratified Sampling

ABSTRACT

This memorandum demonstrates a variance components methodology for partitioning the overall design effect (D) for a ratio mean into stratification (S), unequal weighting (W), and clustering (C) effects, so that D = WSC. In section 2, a sample selection scheme modeled after the National Longitudinal Study of the High School Class of 1972 (NLS) sample is described. In section 3, variance components are defined which can be used to model the variance of an NLS sample total. A second set of components are defined which parallel the appropriate NLS components with the exception that all effects of unequal weighting are removed. That is, components are estimated as if schools were allocated to final strata strictly proportional to the actual student population per stratum; as if the actual student enrollments were used to assign school size measures; and finally, as if there was no nonresponse. In section 4, sample mean squares are presented for estimating the components defined in section 3. Section 4 translates the components definitions developed for sample totals in to the associated components for a domain mean. In the final section, various functions of variance components and sample size parameters are presented for estimating the effects W,S, and C. (Author/CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED175874

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

22U-884-3

VARIANCE COMPONENTS FOR NLS:

PARTITIONING THE DESIGN EFFECT

by

Ralph E. Folsom, Jr.

Statistics Research Division
Research Triangle Institute

TM 007669

January 1975

TABLE OF CONTENTS

		Page
1.	INTRODUCTION	1
2.	NLS SAMPLE DESIGN	3
3.	VARIANCE COMPONENTS MODEL	7
4.	VARIANCE COMPONENTS ESTIMATION	10
	4.1 Regular NLS Component Estimators	10
	4.2 Self-Weighting NLS Component Estimators	12
	4.3 Component Estimators for Per Student Means	14
5.	PARTITIONING THE NLS DESIGN EFFECT	18

1. INTRODUCTION

When one examines the precision of sample statistics from a complex survey design like the NLS, it helps to establish a set of standards for comparison. One standard that comes immediately to mind is the precision that would have been attained had an equal sized simple random sample of respondents been selected. To state this actual variance versus simple random sampling variance comparison, Leslie Kish introduced his "Design Effect" (DEFF), the actual variance of a sample statistic divided by the variance appropriate for an equal sized simple random sample. DEFF measures the combined effect of all the sample selection and estimation procedures which have an impact on the precision of sample statistics. These procedures include:

(1) Stratification;

(2) Unequal weighting attributable to

  (a) Disproportionate stratum allocations (oversampling),

  (b) PPS selections with inexact size measures, and

  (c) Nonresponse adjustments;

(3) Clustering;

(4) Special estimation methods such as ratio adjustments to census population totals by age, race, sex, etc.

The purpose of this memorandum is to demonstrate a variance components methodology for partitioning the overall design effect D for a ratio mean into stratification (S), unequal weighting (W), and clustering (C) effects, so that D = WSC. In section 2, a sample selection scheme modeled after the

NLS sample is described. In section 3, variance components are defined which can be used to model the variance of an NLS sample total. A second set of components are defined which parallel the appropriate NLS components with the exception that all effects of unequal weighting are removed. That is, components are estimated as if schools were allocated to final strata strictly proportional to the actual student population per stratum, as if the actual student enrollments were used to assign school size measures, and finally, as if there was no nonresponse. In section 4, sample mean squares are presented for estimating the components defined in section 3. Section 4 translates the components definitions developed for sample totals into the associated components for a domain mean. In the final section, various functions of variance components and sample size parameters are presented for estimating the effects W, S, and C.

2. NLS SAMPLE DESIGN

Let Y(hij) denote a variate value for the j-tn eligible student in school-i of final stratum-h. The population total of the Y variates is

$$Y(+) = \sum_{h=1}^{H} \sum_{i=1}^{S(h)} \sum_{j=1}^{N(hi)} Y(hij)$$

$$= \sum_{h=1}^{H} \sum_{i=1}^{S(h)} Y(hi)$$

$$= \sum_{h=1}^{H} Y(h) \qquad (1)$$

where H represents the number of final strata (H = 600 for NLS), S(h) represents the number of schools in the stratum-h sampling frame, and N(hi) denotes the number of eligible students in school-hi. To estimate the population total in (1), suppose that a sample of s(+) schools is first allocated to the H final strata, proportional to a stratum size measure A(h) where A(h) may be an estimate of $N(h) = \sum_{i=1}^{S(h)} N(hi)$, the total population of eligibles in stratum-h. These approximate size measures may be adjusted to oversample certain strata, that is

$$A(h) = \begin{cases} k\,\tilde{N}(h) & \text{for } h \in o \\ \tilde{N}(h) & \text{for } h \notin o \end{cases} \qquad (2)$$

where $\tilde{N}(h)$ is the presurvey estimate of eligibles in stratum-h and k is the desired oversampling rate for students in the set of strata denoted by o. The number of schools thusly allocated to stratum-h is

$$s(h) = s(+)\,[A(h)/A(+)] = s(+)\,P(h) \qquad (3)$$

where a plus sign replacing a subscript indicates summation over the range of that subscript. We will assume further, that the H strata were formed by aggregating schools with approximate size measures A(hi) so that the allocations in (3) are essentially integral.

Having assigned schools to final strata, we assume that s(h) schools are drawn from stratum-h with replacement using selection probabilities

$$P(i|h) = [A(hi)/A(h)] \tag{4}$$

at each draw. This with replacement selection model will simplify the variance components partitioning and should provide a reasonable approximation for the NLS without replacement school selections. From each of the s(+) sample schools, we assume that n eligible students are selected with equal probabilities

$$P(j|hi) = 1/N(hi) \tag{5}$$

and with replacement. Nonresponding schools and students will be treated as if they were removed from the sample at random; that is, without regard to their performance. This assumption leads to a nonresponse imputation procedure which replaces missing student data with the average response from participating students in the same school. Similarly, nonresponding schools have their data replaced by the weighted average of cooperating school data from the same stratum. While NLS uses a weighting class approach to student nonresponse imputation, the impact on weight variability should be similar to the scheme described above.

In the following section, a linear identity relating the separate draw

statistics

$$y(hij) = Y(hij)/P(h)\ P(i|h)\ P(j|hi)$$

$$= N(hi)\ Y(hij)/P(hi) \qquad (6)$$

to student, school and stratum effects will be exploited to develop a variance components partitioning for the total estimator

$$\hat{Y} = \bar{y} = \sum_{h=1}^{H} P(h) \sum_{i=1}^{s(h)} \sum_{j=1}^{n(hi)} y(hij)/s(h)n(hi)$$

$$= \sum_{h=1}^{H} P(h) \sum_{i-1}^{s(h)} \bar{y}(hi)/s(h)$$

$$= \sum_{h=1}^{H} P(h)\ \bar{y}(h) = \sum_{h=1}^{H} s(h)\bar{y}(h)/s(+) \quad . \qquad (7)$$

Components will also be developed for a self-weighting version of the NLS sample. This version will differ from the actual NLS sample in the following respects:

(1) The allocation of schools to strata will be proportional to the actual number attending school in the stratum. That is,

$$s(h) = s(+)\ [N(h)/N(+)] = s(+)\ \theta(h) \quad ;$$

(2) The selection of schools within strata is strictly proportional to each school's eligible enrollment. That is, schools are selected with replacement with probabilities

$$\theta(i|h) = [N(hi)/N(h)]$$

at each draw;

(3) Each sample school provides n eligible respondents selected with replacement and with equal probabilities

$$\theta(j|hi) = 1/N(hi) \quad .$$

There is no nonresponse at either the school or student level.

3. VARIANCE COMPONENTS MODEL

To specify variance components for the sample designs and population total estimators presented in section 2, it will be helpful to develop a model identity for the separate draw statistics y(hij). The identity states that

$$y(hij) = \mu + \alpha(h) + \beta(hi) + \varepsilon(hij) \tag{8}$$

where $\mu = Y(+)$ is the population total

$\alpha(h) = \left\{ \frac{Y(h)}{P(h)} - Y(+) \right\}$ is the "effect" of stratum-h.

$\beta(hi) = \left\{ \frac{Y(hi)}{P(h)P(i|h)} - \frac{Y(h)}{P(h)} \right\}$ is the "effect" of school-i within stratum-h.

$\varepsilon(hij) = \left\{ y(hij) - \frac{Y(hi)}{P(h)P(i|h)} \right\}$ is the "effect" of student-j within school-hi.

The student and school effects in (8) are random effects which have zero expectations over all possible school and student samples. The sampling variance of each $\varepsilon(hij)$ from stratum-h is

$$\mathrm{Var}\{\varepsilon(hij)\} = E\{\varepsilon^2(hij)\} = \sum_{i=1}^{S(h)} P(i|h) \sum_{j=1}^{N(hi)} \varepsilon^2(hij)/N(hi)$$

$$= \sum_{i=1}^{S(h)} P(i|h)\, \sigma_2^2(hi)$$

$$= \sigma_2^2(h) \quad . \tag{9}$$

The sampling variance of each school effect from stratum-h is

$$\mathrm{Var}\{\beta(hi)\} = E\{\beta^2(hi)\} = \sum_{i=1}^{S(h)} P(i|h)\beta^2(hi)$$

$$= \sigma_1^2(h) \quad . \tag{10}$$

The with replacement sampling model results in ε and β effects which are independent among themselves and between each other. That is,

$$E\{\varepsilon(hij)\varepsilon(h'i'j')\} = 0 \quad \text{unless } (hij) = (h'i'j')$$

$$E\{\beta(hi)\beta(h'i')\} = 0 \quad \text{unless } (hi) = (h'i')$$

$$E\{\beta(hi)\varepsilon(h'i'j')\} = 0 \quad \text{for all pairs } (hi), (h'i'j').$$

The fixed stratification effects $\alpha(h)$ are such that

$$\sum_{h=1}^{H} P(h)\alpha(h) = 0 \quad . \tag{11}$$

To assess the effect of this stratification, the following strata variance component will be useful:

$$\sigma_0^2 = \sum_{h=1}^{H} P(h)\alpha^2(h) \tag{12}$$

The school and student level components will also be summarized over strata by

$$\sigma_1^2 = \sum_{h=1}^{H} P(h)\sigma_1^2(h) = \sum_{h=1}^{H}\sum_{i=1}^{S(h)} P(h)P(i\,h)\beta^2(hi)$$

and (13)

$$\sigma_2^2 = \sum_{h=1}^{H} P(h)\sigma_2^2(h) = \sum_{h=1}^{H}\sum_{i=1}^{S(h)}\sum_{j=1}^{N(hi)} P(h)P(i|h)\varepsilon^2(hij)/N(hi) \quad .$$

The corresponding student, school and stratum components for the "self-weighting" NLS design are defined in terms of the equally weighted separate draw statistics

$$y_\theta(hij) = Y(hij)/\theta(h)\theta(i|h)\theta(j|hi)$$

$$= N(hi)Y(hij)/\theta(hi)$$

where

$$\theta(hi) = [N(h)/N(+)][N(hi)/N(h)]$$

$$= [N(hi)/N(+)] \quad .$$

The self-weighting nature of the ideal scheme described in (2) is evidenced by the fact that

$$y_\theta(hij) = N(+)\ Y(hij) \quad . \tag{14}$$

Substituting the θ parameters for the P parameters in the "effect" definitions of equation (8) one obtains the corresponding "self-weighting" components. These components will be distinguished from the regular NLS components by using v_0^2, v_1^2, and v_2^2 to depict the stratum, school, and student level parameters.

4. VARIANCE COMPONENTS ESTIMATION

4.1 Regular Ne. Component Estimators

In this section a set of analysis of variance type mean squares and sums of squares will be proposed for estimating the σ^2 and v^2 components defined in the previous section. Motivation for the proposed $\hat{\sigma}^2$ estimators is provided in Folsom, Bayless, and Shah (1971 Social Statistics Section Proceedings of the American Statistical Association). To simplify the calculation of these mean squares and sums of squares, we choose to work with the weighted sample statistics

$$y_w(hij) = w(hij)\, Y(hij) \tag{15}$$

with

$$w(hij) = \left\{ s(h)P(i|h)[n(hi)/N(hi)] \right\}^{-1}$$

$$= N(hi)/s(h)n(hi)\, P(i|h) \quad .$$

The relationship between these weighted sample statistics and the single draw statistics defined in equation (6) is

$$y(hij) = s(h)n(hi)\, y_w(hi\,)/\, P(h) \tag{16}$$

Recalling from (3) that P(h) denotes the achieved school allocation fraction for final stratum-h, one notes that

$$y(hij) = s(+)n(hi)\, y_w(hij) \tag{17}$$

Patterned after the results in Folsom, et. al. (1971) we propose

$$MS_2 = \sum_{h=1}^{H} P(h) \sum_{i=1}^{s(h)} \sum_{j=1}^{n(hi)} [y(hij) - y(hi\cdot)]^2/s(h)[n(hi) - 1]$$

as an estimator for σ_2^2 the student level component. With $P(h) = s(h)/s(+)$, this mean square becomes

$$MS_2 = s(+) \sum_{h=1}^{H} \sum_{i=1}^{s(h)} n^2(hi) \sum_{j=1}^{n(hi)} [y_w(hij) - y_w(hi\cdot)]^2/[n(hi) - 1]$$

$$= s(+) \sum_{h=1}^{H} \sum_{i=1}^{s(h)} n^2(hi)\, MS_2(hi) \qquad (18)$$

in terms of the weighted responses. Another student level mean square which is used as a correction term in the estimation of the school level component is

$$MS_2' = s(+) \sum_{h=1}^{H} \sum_{i=1}^{s(h)} n(hi)\, MS_2(hi) \qquad (19)$$

Two school level mean squares are similarly calculated as

$$MS_1 = s(+) \sum_{h=1}^{H} s(h) \sum_{i=1}^{s(h)} [y_w(hi+) - y_w(h\cdot+)]^2/[s(h) - 1]$$

$$= s(+) \sum_{h=1}^{H} s(h)\, MS_1(h) \qquad (20)$$

$$\text{with } y_w(h\cdot+) = \sum_{i=1}^{s(h)} y_w(hi+)/s(h)$$

and

$$MS_1' = s(+) \sum_{h=1}^{H} MS_1(h) \qquad (21)$$

The between stratum mean square is

$$MS_0 = \sum_{h=1}^{H} P(h) \left\{ \frac{y_w(h++)}{P(h)} - y_w(+++) \right\}^2$$

$$= s(+) \sum_{h=1}^{H} [y_w^2(h++)/s(h)] - y_w^2(+++) \qquad (22)$$

Combining the five mean squares in equations (18) through (22) one can form the following unbiased variance components (assuming the with replacement selection scheme):

$$\hat{\sigma}_2^2 = MS_2 \tag{23}$$

$$\hat{\sigma}_1^2 = (MS_1 - MS_2') \tag{24}$$

$$\hat{\sigma}_0^2 = [MS_0 - MS_1' + MS_1/s(+)] \quad . \tag{25}$$

4.2 Self-Weighting NLS Component Estimators

To estimate the v^2 components, the sample data can be used to estimate the θ parameters, true relative size measures, which distinguish these components from the σ^2 values. Specifically, we use

$$\hat{\theta}(i|h) = w(hi+)/w(h++)$$

and (26)

$$\hat{\theta}(h) = w(h++)/w(+++) \quad .$$

with the plus signs replacing weight subscripts which have been summed over. The following mean squares and sums of squares yield estimates of the v^2 components which would be unbiased if the ratio estimates in (26) were in fact the true θ values free of sampling errors:

$$ms_2 = w(+++) \sum_{h=1}^{H} \sum_{i=1}^{s(h)} n^2(hi)\, MS_2(hi)/w(hi+, \tag{27}$$

$$ms_2' = w(+++) \sum_{h=1}^{H} \sum_{i=1}^{s(h)} n(hi)\, MS_2(hi)/w(hi+) \tag{28}$$

$$SS_2 = w(+++) \sum_{h=1}^{H} \sum_{i=1}^{s(h)} \sum_{j=1}^{n(hi)} y_w^2(hij)/w(hij) \tag{29}$$

$$ms_1 = w(+++) \sum_{h=1}^{H} s(h)\, MS_1(h)/w(h++) \tag{30}$$

$$SS_1 = w(+++) \sum_{h=1}^{H} \sum_{i=1}^{s(h)} y_w^2(hi+)/w(hi+) \tag{31}$$

$$SS_0 = w(+++) \sum_{h=1}^{H} y_w^2(h++)/w(h++) \quad . \tag{32}$$

The associated variance components for the self-weighting version of NLS are:

$$\hat{v}_2^2 = ms_2 \tag{33}$$

$$\hat{v}_2^2 = (SS_1 - SS_0 + ms_1 - ms_2') \tag{34}$$

$$\hat{v}_0^2 = [SS_0 - y_w^2(+++) + MS_1/s(+) - ms_1] \quad . \tag{35}$$

Noticing that

$$(SS_2 - SS_1) = ms_2 - ms_2'$$

one can also specify $\hat{v}_2^2$ as

$$\hat{v}_2^2 = (SS_2 - SS_1 + ms_2') \quad . \tag{36}$$

Using the version of $\hat{v}_2^2$ in (36), one finds that

$$\hat{v}_0^2 + \hat{v}_1^2 + \hat{v}_2^2 = [SS_2 - y_w^2(+++) + MS_1/s(+)] \quad . \tag{37}$$

Recalling that

$$SS_2 = w(+++) \sum_{h=1}^{H} \sum_{i=1}^{s(h)} y_w^2(hij)/w(hij)$$

$$= w(+++) \sum_{h=1}^{H} \sum_{i=1}^{s(h)} w(hij)\, Y^2(hij)$$

we see that

$$E\{SS_2\} \doteq N(+) \sum_{h=1}^{H} \sum_{i=1}^{s(h)} Y^2(hij) \quad .$$

It is also clear, that

$$E\{y_w^2(+++)\} = Y^2(+++) + Var\{y_w(+++)\} \quad .$$

Since $MS_1/s(+)$ is an unbiased estimate for $Var\{y_w(+++)\}$, assuming with replacement selection, we see that $\hat{v}_0^2 + \hat{v}_1^2 + \hat{v}_2^2$ is a reasonably unbiased estimate for the simple random sampling variance of $\hat{Y}(+++)$. Since the relative bias of $y_w^2(+++)$ taken as an estimator of $Y^2(+++)$ is $Var[y_w(+++)]/Y^2(+++)$, or the rel-variance of $y_w(+++)$, one could probably get by with using

$$\tilde{v}_0^2 = [SS_0 - y_w^2(+++) - ms_1] \tag{38}$$

which leads to the more familiar SRS variance component estimator

$$\tilde{v}^2 + \hat{v}^2 + \hat{v}_2^2 = [SS_2 - y_w^2(+++)] \quad . \tag{39}$$

4.3 Component Estimates for Per Student Means

To form the variance components for a mean per student in domain D; namely,

$$\hat{\bar{Y}}_D = \left\{ \frac{\sum_{h=1}^{H} \sum_{i=1}^{S(h)} \sum_{j=1}^{n(hi)} w(hij)\, X_D(hij)\, Y(hij)}{\sum_{h=1}^{H} \sum_{i=1}^{S(h)} \sum_{j=1}^{n(hi)} w(hij)\, X_D(hij)} \right\}$$

$$= (\hat{Y}_D/\hat{X}_D) \quad . \tag{40}$$

where $X_D(hij)$ is a zero-one indicator variable for domain D, one can substitute the linearized variate

$$x_D(hij) = X_D(hij)\,[Y(hij) - \bar{Y}_D]/X_D \tag{41}$$

for Y(hij) in the formulae for the σ^2 and v^2 components, where $\bar{Y}_D$ is the domain mean (population parameter) and X_D is the domain size. In the estimation formulae one substitutes

$$\hat{Z}_D(hij) = X_D(hij)\ [Y(hij) - \hat{\bar{Y}}_D]/\hat{X}_D \tag{42}$$

for Y(hij), or $z_D(hij) = w(hij)\ \hat{Z}(hij)$ for $y_w(hij)$.

For example, σ_0^2 becomes

$$\sigma_0^2 = \sum_{h=1}^{H} Z_D^2(h)/P(h) ,$$

since $Z_D(+) = 0$. Now,

$$Z_D(h) = [Y_D(h) - \bar{Y}_D\ X_D(h)]/X_D$$

$$= \theta_D(h)\ [\bar{Y}_D(h) - \bar{Y}_D]$$

where $\theta_D(h) = [X_D(h)/X_D]$ is the fraction of the domain population in stratum-h. Then

$$\sigma_0^2 = \sum_{h=1}^{H} \left[\frac{\theta_D^2(h)}{P(h)}\right]\ [\bar{Y}_D(h) - \bar{Y}_D]^2 \quad . \tag{43}$$

Similarly,

$$v_0^2 = \sum_{h=1}^{H} \left[\frac{\theta_D^2(h)}{\theta(h)}\right]\ [\bar{Y}_D(h) - \bar{Y}_D]^2 \quad . \tag{44}$$

The school components take the form

$$\sigma_1^2 = \sum_{h=1}^{H} \sum_{i=1}^{S(h)} P(hi) \left[\left\{\left[\frac{\theta_D(hi)}{P(hi)}\right] \bar{Y}_D(hi) - \left[\frac{\theta_D(h)}{P(h)}\right] \bar{Y}_D(h)\right\}\right.$$

$$\left. - \bar{Y}_D \left\{\left[\frac{\theta_D(hi)}{P(hi)}\right] - \left[\frac{\theta_D(h)}{P(h)}\right]\right\}\right]^2 \quad . \tag{45}$$

In v_1^2, $\theta(hi)$ replaces $P(hi)$. The student level components are of the form

$$\sigma_2^2 = \sum_{h=1}^{H} \sum_{i=1}^{S(h)} \left[\frac{N^2(hi)}{P(hi)X_D^2}\right] \sum_{j=1}^{N(hi)} \left[\left\{Y_D(hij) - \left[\frac{X_D(hi)}{N(hi)}\right] \bar{Y}_D(hi)\right\} - \bar{Y}_D\left\{X_D(hij) - \frac{X_D(hi)}{N(hi)}\right\}\right]^2 \Big/ N(hi) \tag{46}$$

where $Y_D(hij) = X_D(hij)\, Y(hij)$. The corresponding self-weighting student level component v_2^2 has $\theta(hi)$ replacing $P(hi)$.

For the total population average $\bar{Y} = [Y(+)/N]$ one sees that

$$\sigma_0^2 = \sum_{h=1}^{H} \left[\frac{\theta^2(h)}{P(h)}\right] [\bar{Y}(h) - \bar{Y}]^2$$

$$v_0^2 = \sum_{h=1}^{H} \theta(h)\, [\bar{Y}(h) - \bar{Y}]^2 = \sum_{h=1}^{H} N(h)\, [\bar{Y}(h) - \bar{Y}]^2/N$$

$$\sigma_1^2 = \sum_{h=1}^{H} \sum_{i=1}^{S(h)} P(hi) \left[\left\{\left[\frac{\theta(hi)}{P(hi)}\right] \bar{Y}(hi) - \left[\frac{\theta(h)}{P(h)}\right] \bar{Y}(h)\right\} - \bar{Y}\left\{\left[\frac{\theta(hi)}{P(hi)}\right] - \left[\frac{\theta(h)}{P(h)}\right]\right\}\right]^2$$

$$v_1^2 = \sum_{h=1}^{H} \sum_{i=1}^{S(h)} N(hi)\, [\bar{Y}(hi) - \bar{Y}(h)]^2/N$$

$$\sigma_2^2 = \sum_{h=1}^{H} \sum_{i=1}^{S(h)} \left[\frac{\theta^2(hi)}{P(hi)}\right] \sum_{j=1}^{N(hi)} [Y(hij) - \bar{Y}(hi)]^2/N(hi)$$

$$v_2^2 = \sum_{h=1}^{H} \sum_{i=1}^{S(h)} \theta(hi) \sum_{j=1}^{N(hi)} [Y(hij) - \bar{Y}(hi)]^2/N(hi)$$

$$= \sum_{h=1}^{H} \sum_{i=1}^{S(h)} \sum_{j=1}^{N(hi)} [Y(hij) - \bar{Y}(hi)]^2/N$$

Combining v_0^2, v_1^2, and v_2^2 we see that

$$v_0^2 + v_1^2 + v_2^2 = \sum_{h=1}^{H} \sum_{i=1}^{S(h)} \sum_{j=1}^{N(hi)} [Y(hij) - \bar{Y}]^2/N$$

the simple random sampling variance component for the population mean. If one selected students with replacement with probabilities $P(hij) = P(hi)/N(hi)$, the sampling variance component would be

$$\sigma_0^2 + \sigma_1^2 + \sigma_2^2 = \sum_{h=1}^{H} \sum_{i=1}^{S(h)} \left[\frac{\theta^2(hi)}{P(hi)}\right] \sum_{j=1}^{N(hi)} [Y(hij) - \bar{Y}]^2/N(hi) \quad . \qquad (47)$$

If the within school variance components

$$\Sigma^2(hi) = \sum_{j=1}^{N(hi)} [Y(hij) - \bar{Y}]^2/N(hi) \doteq \Sigma^2$$

for all the schools, then

$$W_r = \left\{ \frac{\sigma_0^2 + \sigma_1^2 + \sigma_2^2}{v_0^2 + v_1^2 + v_2^2} \right\} \doteq \sum_{h=1}^{H} \sum_{i=1}^{S(h)} \left[\frac{\theta^2(hi)}{P(hi)}\right] \qquad (48)$$

represents the increase in variance of the single stage unequal probability selection over the single stage equal probability selection.

5. PARTITIONING THE NLS DESIGN EFFECT

A more appropriate measure of the effect of unequal weighting contrasts the variance of NLS's stratified two-stage cluster sample with disproportionate allocations to strata, imprecise school size measures, and nonresponse against the corresponding proportionally allocated sample with school selection probabilities strictly proportional to actual enrollments and no nonresponse. The NLS variance model is

$$\Sigma_1^2 = \frac{\sigma_1^2}{s} + \frac{\sigma_2^2}{sn} \tag{49}$$

using s for s(+), the total number of schools selected, and assuming n(hi) = n for all schools-hi.

The self-weighting model is

$$V_1^2 = \frac{v_1^2}{s} + \frac{v_2^2}{sn} \quad . \tag{50}$$

Therefore

$$W = \Sigma_1^2 / V_1^2$$

is the increase in variance attributable to oversampling, imprecise size measures, and nonresponse. After removing the effect of unequal weighting, one can examine the impact of school stratification by contrasting

$$V_2^2 = \frac{(v_0^2 + v_1^2)}{s} + \frac{v_2^2}{sn} \tag{51}$$

with V_1^2, since V_2^2 models the variance of an unstratified selection of s schools with probabilities proportional to actual enrollment and an equal probability selection of n students per school. The ratio

$$S = V_1^2 / V_2^2 \tag{52}$$

represents the reduction in total variance that can be achieved by stratifying schools after removing the effect of unequal weighting.

The effect of clustering the sample of students in schools after removing the effects of unequal weighting and stratification can be measured by contrasting V_2^2 against the simple random sampling model

$$V_3^2 = (v_0^2 + v_1^2 + v_2^2)/sn \tag{53}$$

This effect of clustering is

$$C = (V_2^2/V_3^2) = [1 + (n - 1)\,\delta_c] \tag{54}$$

where

$$\delta_c = (v_0^2 + v_1^2)/(v_0^2 + v_1^2 + v_2^2)$$

is the intra-school cluster correlation for an unstratified sample of schools. If $\delta_{c/s}$ denotes the intra-school cluster correlation for a stratified sample of schools, where

$$\delta_{c/s} = v_1^2/(v_1^2 + v_2^2) \tag{55}$$

and S_r represents the reduction in variance for a stratified random sample with proportional allocation over a simple random sample, then S, the total effect of stratification becomes

$$S = S_r \left\{ \frac{1 + (n - 1)\,\delta_{c/s}}{1 + (n - 1)\,\delta_c} \right\} = S_r\, S_c \tag{56}$$

where

$$S_r = \left(\frac{v_1^2 + v_2^2}{v_0^2 + v_1^2 + v_2^2} \right) .$$

The effect of unequal weighting can also be partitioned into a stratified random sampling effect

$$W_{sr} = (\sigma_1^2 + \sigma_2^2)/(v_1^2 + v_2^2) \tag{57}$$

and an additional effect to account for stratified cluster sampling; namely,

$$W_{sc} = \left\{ \frac{1 + (n - 1)\, \delta_{c/sw}}{1 + (n - 1)\, \delta_{c/s}} \right\} \tag{58}$$

where

$$\delta_{c/sw} = \sigma_1^2/(\sigma_1^2 + \sigma_2^2) \quad .$$

Comparing the expressions for σ_1^2 and σ_2^2, one might expect the unequal weighting to have a greater impact at the student level; that is, we would anticipate that

$$(\sigma_2^2/v_2^2) > (\sigma_1^2/v_1^2) \quad .$$

If this conjecture is true for the NLS design, then one would find that

$$\delta_{c/sw} < \delta_{c/s}$$

which implies that $W_{sc} < 1$. This leads one to expect that the total effect of unequal weighting

$$W = W_{sr}\, W_{sc} \tag{59}$$

will be smaller than W_{sr}. One could further partition W_{sr} into W_r, where

$$W_r = \left(\frac{\sigma_0^2 + \sigma_1^2 + \sigma_2^2}{v_0^2 + v_1^2 + v_2^2} \right)$$

represents the weighting effect on a single stage unstratified selection

previously introduced in (48). Recall that S_r was the reduction in variance of a stratified random sample with proportional allocation over a simple random sample. If S_{rw} represents the amount of reduction for a stratified random sample with NLS's disproportionate allocation, then

$$W_{sr} = W_r (S_{rw}/S_r) = W_r W_s \quad . \tag{60}$$

Since the disproportionate allocation would be expected to reduce the effectiveness of NLS stratification, at least for total population statistics, one would expect W_s to be greater than 1.

Combining these effects of unequal weighting, stratification, and clustering, the following model for the total design effect can be specified:

$$D = \Sigma_1^2/V_3^2 = WSC \quad . \tag{61}$$

With further partitioning of the W, S and C effects one obtains

$$D = W_r W_s W_{sc} S_r S_c C \tag{62}$$

where

$$W_r = \left(\frac{\sigma_0^2 + \sigma_1^2 + \sigma_2^2}{v_0^2 + v_1^2 + v_2^2}\right) \equiv$$ The effect of unequal selection probabilities on an unstratified single stage sample of students.

$$W_s = (S_{rw}/S_r) \equiv$$ The reduction in the effectiveness of stratification with disproportionate allocation and unequal probability selection.

$$S_{rw} = \left(\frac{\sigma_1^2 + \sigma_2^2}{\sigma_0^2 + \sigma_1^2 + \sigma_2^2}\right) \equiv$$ The effect of stratification on a single stage selection of students with disproportionate stratum allocation and unequal probability selection.

$$S_r = \left(\frac{v_1^2 + v_2^2}{v_0^2 + v_1^2 + v_2^2} \right) \equiv$$ The effect of stratification on a single stage equal probability selection of students.

$$W_{sc} = \left\{ \frac{1 + (n - 1)\, \delta_{c/sw}}{1 + (n - 1)\, \delta_{c/s}} \right\} \equiv$$ The impact of unequal weighting on the effect of clustering in a stratified selection of schools and students.

$$\delta_{c/s} = \left(\frac{v_1^2}{v_1^2 + v_2^2} \right) \equiv$$ The intra-school cluster correlation which contains the effect of stratification, but is free of unequal weights.

$$S_c = \left\{ \frac{1 + (n - 1)\, \delta_{c/s}}{1 + (n - 1)\, \delta_c} \right\} \equiv$$ The impact of stratification on the effect of clustering in a self-weighting selection of schools and students.

$$\delta_c = \left(\frac{v_0^2 + v_1^2}{v_0^2 + v_1^2 + v_2^2} \right) \equiv$$ The intra-school cluster correlation from an unstratified self-weighting selection of schools and students.

$$C = [1 + (n - 1)\, \delta_c] \equiv$$ The effect of clustering in an unstratified self-weighting selection of schools and students.